1351.0.55.060 - Research Paper: Personal Income Tax and Migrants Integrated Dataset (PITMID) 2011-12 Quality Assessment, Oct 2016
Latest ISSUE Released at 11:30 AM (CANBERRA TIME) 19/10/2016 First Issue

Summary

Executive Summary

EXECUTIVE SUMMARY

The Australian Bureau of Statistics created the Personal Income Tax and Migrants Integrated Dataset (PITMID) by linking the Australian Taxation Office Personal Income Tax (PIT) records with migrant records from the Australian Government's Settlement Database (SDB). The PITMID Project initially began in 2013 with a linking feasibility study. During the study, almost a million migrant settlement records (54%) linked to a PIT record demonstrating that the linking was feasible. The study concluded that the linked 2009–10 and 2010–11 PITMID dataset provides valuable new information on recent permanent and provisional migrant taxpayers' personal income. In 2015, the 2009–10 and 2010–11 PITMID data was released in *Personal Income of Migrants, Australia, Experimental* (ABS cat. no. 3418.0).

PITMID contains key personal income variables (employee income, own unincorporated business income, investment income, other income and foreign income) and SDB variables (visa subclass, application status (primary or secondary), location (onshore or offshore), country of birth and year of arrival for Skill, Family, Humanitarian, Other permanent and provisional visa holders). The SDB records are linked to the PIT records using variables such as name, date of birth and address. Relevant legislation and guidelines, including the *Privacy Act 1988* and the *High Level Principles for Data Integration Involving Commonwealth Data for Statistical and Research Purposes* were adhered to, protecting the privacy of individuals on both datasets.

This PITMID study was conducted to assess the effects of the change in the linking methodology introduced in 2016 for the 2011–12 PITMID linkage. The 2009–10 and 2010–11 PITMID linkage employed a combined deterministic and probabilistic linking methodology. The new linking methodology utilises a Statistical Analysis Software (SAS) macro known as the Deterministic linking Macro (D-MAC) for a purely deterministic approach. The D-MAC links two datasets using a simple set of rules and then outputs linked record pairs with a calculated measure of accuracy. The study briefly outlines the original and new linking methodologies and presents the results of the analyses conducted to assess the quality of the 2011–12 PITMID linkage compared with the 2009–10 and 2010–11 PITMID linkages. This was done by running the D-MAC over the full SDB dataset and the 2009–10 and 2010–11 PIT datasets.

The new methodology utilising the D-MAC was found to be much quicker to administer and produced high quality results, while enabling comparison between the annual series. The linking results generated by D-MAC showed almost 95% of the SDB records either linked to the same PIT record (as the previous linking) or did not link to a PIT record. For this reason,

the links generated for 2009–10 and 2010–11 were retained for the 2011–12 linkage process. It is anticipated that the PITMID Project will continue to use the D-MAC for linking in future. The D-MAC is also becoming the preferred linking method for other important ABS data integration projects. Utilisation of the same linking methodology for PITMID will ensure that the project is well placed should any further opportunities arise for linking with other datasets in the future.

About this Release

The Australian Bureau of Statistics created the Personal Income Tax and Migrants Integrated Dataset (PITMID) by linking Australian Taxation Office personal income tax records with migrant records from the Australian Government's Settlement Database (SDB). In 2015, PITMID data for 2009–10 and 2010–11 was released in Personal Income of Migrants, Australia, Experimental (ABS cat. no. 3418.0). In 2016, this PITMID study was conducted to assess the effects of a change in the linking methodology for the 2011–12 PITMID. The study briefly outlines the original and new linking methodologies and presents the results of the analyses conducted to assess the quality of the 2011–12 PITMID linkage compared with the 2009–10 and 2010–11 PITMID linkages. The new deterministic record linkage methodology was found to be much quicker to administer and produced high quality results, while enabling comparison between the annual series.